

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

2

REPORT DOCUMENTATION PAGE

AD-A197 748

IC
CTE

1b. RESTRICTIVE MARKINGS

3. DISTRIBUTION/AVAILABILITY OF REPORT

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

AUG 18 1988

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

Technical Report #3

5. MONITORING ORGANIZATION REPORT NUMBER(S)

AFOSR-TR- 88-0775

6a. NAME OF PERFORMING ORGANIZATION

Texas A&M University

6b. OFFICE SYMBOL
(If applicable)

7a. NAME OF MONITORING ORGANIZATION

Air Force Office of Scientific Research

6c. ADDRESS (City, State and ZIP Code)

College Station, TX 77853

7b. ADDRESS (City, State and ZIP Code)

Same as Abo 8C

8a. NAME OF FUNDING/SPONSORING
ORGANIZATION

AFOSR

8b. OFFICE SYMBOL
(If applicable)

NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER

AFOSR-87-49620-85-C-0144

8c. ADDRESS (City, State and ZIP Code)

Bldg 410
Bolling Air Force Base
Washington, DC 20332

10. SOURCE OF FUNDING NOS.

PROGRAM
ELEMENT NO.PROJECT
NO.TASK
NO.WORK UNIT
NO.

61102F

2304

176

11. TITLE (Include Security Classification)

OPTIMAL RATES OF CONVERGENCE FOR DECONVOLVING A DENSITY

12. PERSONAL AUTHOR(S)

Carroll, Raymond J. and Hall, Peter

13a. TYPE OF REPORT

Technical Journal

13b. TIME COVERED

FROM 8/87 TO 8/88

14. DATE OF REPORT (Yr., Mo., Day)

15. PAGE COUNT

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD GROUP SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

deconvolution, density estimation, errors, in variables,
measurement error, rates, of convergence.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Suppose we observe the sum of two independent random variables X and Z , where Z denotes measurement error and has a known distribution, and where the unknown density f of X is to be estimated. It is shown that if Z is normally distributed and if f has k bounded derivatives, then the fastest attainable convergence rate of any nonparametric estimator of f is only $(\log n)^{-k/2}$. Therefore deconvolution with normal errors may not be a practical proposition. Other error distributions are also treated. Stefanski-Carroll (1987b) estimators achieve the optimal rates. Our results have versions for multiplicative errors, where they imply that even optimal rates are exceptionally slow.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT

UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION

UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL

Major Brian Woodruff

22b. TELEPHONE NUMBER
(Include Area Code)

(202) 767-5026

22c. OFFICE SYMBOL

NM

AFOSR-TR- 88 - 0775

Optimal rates of convergence
for deconvolving a density

Technical Report #3

Raymond J. Carroll and Peter Hall

Approved For	
Distribution Category	
Classification	
Declassification	
By	
Date/Action/	
Availability Codes	
Dist	Avail and/or Special
A-1	



8

Optimal rates of convergence for deconvolving a density

Raymond J. Carroll

Department of Statistics
Texas A & M University
College Station, TX 77843 (USA)

Peter Hall

Department of Statistics
Australian National University
Canberra ACT 2601 (Australia)

SUMMARY. Suppose we observe the sum of two independent random variables X and Z , where Z denotes measurement error and has a known distribution, and where the unknown density f of X is to be estimated. It is shown that if Z is normally distributed and if f has k bounded derivatives, then the fastest attainable convergence rate of any nonparametric estimator of f is only $(\log n)^{-k/2}$. Therefore deconvolution with normal errors may not be a practical proposition. Other error distributions are also treated. Stefanski-Carroll (1987b) estimators achieve the optimal rates. Our results have versions for multiplicative errors, where they imply that even optimal rates are exceptionally slow.

KEY WORDS. Deconvolution; density estimation; errors-in-variables; measurement error; rates of convergence.

1. INTRODUCTION

Suppose we wish to gain information about the density f of a random variable X , but because of measurement error can only observe $Y = X + Z$, where the measurement error Z is independent of X . Assume Z has a known density function f_Z with characteristic function ϕ_Z . Our paper addresses the question: from a sample Y_1, \dots, Y_n , how well can f be estimated?

Applied problems in which knowledge of f is required are discussed by Mendelsohn & Rice (1983), see also Medgyessy (1977). Nonparametric estimates of f are discussed by Stefanski & Carroll (1987b).

An application of our results is to the nonparametric Empirical Bayes problem, see Maritz (1970) and Berger (1980). Here f represents the prior distribution for a sequence of location parameters X_1, \dots, X_n . The idea is to estimate the prior nonparametrically, as opposed to the alternative device of specifying a parametric form for the prior with parameters to be estimated. Our paper addresses the question: how well can a prior be estimated nonparametrically?

Another application is to the problem of measurement error models (errors-in-variables) for nonlinear regression and generalized linear models; see Stefanski & Carroll (1987a). Other recent papers include Carroll *et al.* (1984), Stefanski & Carroll (1985), Stefanski (1985) and Schafer (1987). In this problem, X is the true predictor but because of measurement error Z we can observe only $Y = X + Z$. While the middle two references use a sensitivity analysis approach, Carroll *et al.* (1984) and Schafer (1987) assume a specific distributional form for f . Our paper addresses the question of how well the data can be used in a nonparametric way to suggest a parametric form for f . Schafer (1987) shows that in generalized linear models, the EM algorithm for maximum likelihood requires knowledge of the first two conditional moments of X given Y and the response variable in the generalized linear model. Other problems will require the conditional moments of X given Y . In either case, how well these conditional moments can be estimated from data depends

crucially on how well f can be estimated from data.

The case of normal measurement error is particularly important. We show in this paper that if f has k bounded derivatives, and if errors are normal, then the fastest rate of convergence of any estimator of f is only $(\log n)^{-k/2}$, and that this rate is achieved by a kernel estimator of Stefanski-Carroll (1987b) type. This very slow rate suggests that deconvolution may not be a practical procedure with normal errors, even if optimal estimators are employed. With $k = 2$, it also follows that the best achievable rate for estimating the distribution function of X can be no faster than $(\log n)^{-3/2}$. Thus, even estimating probabilities for X is difficult.

We also show that Stefanski-Carroll estimators attain optimal convergence rates for many other error distributions, such as gamma, exponential and double exponential. For example, the optimal achievable rate in the double exponential case is $n^{-k/(2k+5)}$. Our results indicate that if the error density is compactly supported and infinitely differentiable then the optimal convergence rate is slower than n^{-a} for any $a > 0$. Deconvolving a density with smooth measurement error is intrinsically difficult, with convergence rates much slower than those usually encountered in density estimation.

Our results have obvious implications for models with multiplicative error, $Y = XZ$, which may be expressed additively by taking logs. The density of $\log Z$ is infinitely differentiable in many important cases, such as when Z is gamma or lognormal, and so convergence rates are extremely slow. Hence, deconvolution is difficult when errors are multiplicative.

Of course, our lower bounds to convergence rates continue to apply when error distributions are known imperfectly, for example when errors are normal with unknown variance. In such cases, where the error distribution is specified up to estimable parameters, the distribution can often be estimated $n^{\frac{1}{2}}$ -consistently by replication. Since estimators of the X -density f converge at rates considerably slower than $n^{-\frac{1}{2}}$, replacing the true error distribution by its estimated version does not measurably affect convergence rates

of Stefanski-Carroll estimators. Hence, both our lower and upper bounds to convergence rates apply when error distributions are imperfectly specified, up to a parametric form.

The next section gives details of our calculations in the case of normal measurement errors. In section 3 we briefly discuss other error distributions.

2. DECONVOLUTION WHEN ERRORS ARE NORMAL

Write $C_k(B)$ for the class of k -times differentiable densities f having $\sup f \leq B$ and $\sup |f^{(k)}| \leq B$. Let X have density f , Z be normal $N(0,1)$ independent of X , and $Y = X + Z$. The following theorem provides bounds to the accuracy with which $f \in C_k(B)$ can be estimated from an n -sample of Y 's.

Let x_0 be any real number, and $\hat{f}(x_0)$ be any nonparametric estimator of $f(x_0)$, based on an n -sample of Y 's.

Theorem 1. *Assume that the error distribution is normal $N(0,1)$. If, for some sequence of positive constants $\{a_n, n \geq 1\}$, we have*

$$\liminf_{n \rightarrow \infty} \inf_{f \in C_k(B)} P_f \{ |\hat{f}(x_0) - f(x_0)| \leq a_n \} = 1 \quad (2.1)$$

for each $B > 0$, then

$$\lim_{n \rightarrow \infty} (\log n)^{k/2} a_n = \infty. \quad (2.2)$$

Theorem 1 declares that the rate of convergence of \hat{f} to f cannot be faster than $(\log n)^{-k/2}$, over densities in $C_k(B)$. Kernel estimators attaining this rate of convergence may be constructed as follows; see Stefanski and Carroll (1987b). Let G be a symmetric function vanishing outside $(-1,1)$, having $k+2$ bounded derivatives on $(-\infty, \infty)$, and satisfying $G(t) = 1 + O(|t|^k)$ as $t \rightarrow 0$. Put $h \equiv (2/\log n)^{1/2}$,

$$G(w, h) \equiv (2\pi)^{-1} \int \cos(tw/h) G(t) \exp\{(t/h)^2/2\} dt$$

and $\hat{f}(x) \equiv (nh)^{-1} \sum_j G(Y_j - x, h)$, where $\{Y_1, \dots, Y_n\}$ is a random sample from the distribution of Y . We have the following converse to Theorem 1.

Theorem 2. Assume that the error distribution is normal $N(0, 1)$. If the constants a_n satisfy (2.2), and if \hat{f} is the kernel estimator just defined, then (2.1) holds for each real number x_0 and each $B > 0$.

Theorem 2 is easily proved, as follows. Put $K(y) \equiv (2\pi)^{-1} \int e^{i\lambda y} G(t) dt$, a real-valued function integrating to unity. Integrating by parts $k + 2$ times we see that $|K(y)| \leq C(1 + |y|^{k+2})^{-1}$. By Fourier inversion, $\int y^j K(y) dy = 0$ for $1 \leq j \leq k - 1$, implying that K is a k 'th-order kernel (Prakasa Rao 1983, p.42). If $Y = X + Z$ then $E\{G(Y - x, h) | X\} = K\{(X - x)/h\}$, so that the mean of $\hat{f}(x)$ is the same as that of a classical k 'th-order kernel estimator based on an n -sample from the X -population. Therefore $|E\hat{f}(x) - f(x)| \leq C_1(B)h^k$ uniformly in $f \in C_k(B)$ (Prakasa Rao 1983, p.47). Furthermore,

$$\begin{aligned} nh^2 \text{var}\{\hat{f}(x)\} &\leq E\{G(Y - x, h)^2\} = E[E\{G(Y - x, h)^2 | X\}] \\ &\leq C_2(B) \iint_{|s|, |t| \leq 1} \exp[(2h^2)^{-1}\{s^2 + t^2 - (s + t)^2\}] ds dt \\ &\leq 4C_2(B) \int_0^1 \int_0^1 \exp(st/h^2) ds dt \leq 4C_2(B) \exp(1/h^2), \end{aligned}$$

whence, noting that $h = (2/\log n)^{1/2}$,

$$\begin{aligned} \sup_{f \in C_k(B)} P_f\{|\hat{f}(x) - f(x)| > a_n\} &\leq a_n^{-2} \sup_{f \in C_k(B)} \{\text{var} \hat{f}(x) + |E\hat{f}(x) - f(x)|^2\} \\ &\leq C_3(B) a_n^{-2} \{(nh^2)^{-1} e^{1/h^2} + h^{2k}\} \rightarrow 0. \end{aligned}$$

This proves Theorem 2.

Finally we derive Theorem 1. To simplify notation we relocate so that $x_0 = 0$, and rescale so that Z is normal $N(0, \frac{1}{2})$, with density $\psi(z) \equiv \pi^{-1/2} e^{-z^2}$. Let $\sigma \geq 1$, and write f_0 for the $N(0, \sigma^2)$ density; l for the integer part of $\log n$; $b_j \equiv 2^{-j} \{(2j)!\}^{-1/2} j^{1/4}$; $\eta \equiv l^{-k/2} \epsilon^k \delta B$, where $\epsilon, \delta \in (0, \frac{1}{2}]$ are fixed; and H_0, H_1, \dots for Hermite polynomials orthogonal with respect to ψ . The following properties are obtainable from Magnus *et al.* (1966, p.252) and Sansone (1959, p.324): $H_j(-x) = (-1)^j H_j(x)$;

$$\exp\{2x\epsilon y - (\epsilon y)^2\} = \sum_{j=0}^{\infty} H_j(x) (\epsilon y)^j / j!; \quad (2.3)$$

$$\int H_i(x) H_j(x) e^{-x^2} dx = \pi^{\frac{1}{2}} 2^i i! \quad \text{if } i = j, 0 \text{ otherwise;} \quad (2.4)$$

$$\int H_{2i}(x) x^{2j} \psi(x) dx = (2j)! / \{4^{j-i} (j-i)!\}; \quad (2.5)$$

$$|b_j H_{2j}(x) \psi(x)| \leq C(1 + |x|^{5/2}) e^{-x^2/2}; \quad (2.6)$$

$$\eta \sup |(d/dx)^k b_l H_{2l}(x/\epsilon) \psi(x/\epsilon)| \leq C\delta B, \quad (2.7)$$

where C depends only on k .

Put $f_n(x) \equiv f_0(x) + \eta b_l H_{2l}(x/\epsilon) \psi(x/\epsilon)$. By (2.6), and since $\eta(n) \rightarrow 0$ and $\epsilon < 1 \leq \sigma$, f_n is a density for large n . If X has density f_0 or f_n then $Y = X + Z$ has density g_0 or g_n respectively, where g_0 is the $N(0, \sigma^2 + \frac{1}{2})$ density, $g_n(x) \equiv g_0(x) + \eta b_l h_l(x)$ and

$$h_l(x) \equiv \int H_{2l}(y/\epsilon) \psi(y/\epsilon) \psi(x-y) dy = \epsilon \psi(x) \sum_{j=l}^{\infty} H_{2j}(x) \epsilon^{2j} \{4^{j-l} (j-l)!\}^{-1},$$

using (2.3) and (2.5). Since $\psi(x)^2/g_0(x) \leq C_1 e^{-x^2}$ then

$$\begin{aligned} I &\equiv \int (g_n - g_0)^2 (g_0)^{-1} \leq C_1 (\eta b_l)^2 \int h_l(x)^2 e^{x^2} dx \\ &= C_2 \epsilon^{2k+2} \delta^2 2^{2l} \{(2l)!\}^{-1} l^{\frac{1}{2}-k} \sum_{j=l}^{\infty} (\epsilon^4/4)^j (2j)! \{(j-l)!\}^{-2}, \end{aligned} \quad (2.8)$$

using (2.4). But $\{(2l)!\}^{-1} \leq C_3 (l!)^{-2} 2^{-2l} l^{\frac{1}{2}}$, $(2j)! \leq C_3 (j!)^2 2^{2j} j^{-\frac{1}{2}}$ and $j!/(j-l)! = \binom{j}{l} l! \leq 2^j l!$. Hence, remembering that $\epsilon \leq \frac{1}{2}$,

$$I \leq C_4 \epsilon^{2k+2} \delta^2 l^{1-k} \sum_{j=l}^{\infty} (4\epsilon^4)^j j^{-\frac{1}{2}} \leq C_5 (\epsilon, \delta) l^{\frac{1}{2}-k} (4\epsilon^4)^l = o(n^{-1}). \quad (2.9)$$

Given $B > 0$, we see from (2.6) and (2.7) that by choosing σ large and δ small, not depending on B , we may ensure that $f_0, f_n \in C_k(B)$ for large n . For an event A , let $P_n(A)$ and $P_0(A)$ denote the probability of A under f_n and f_0 respectively. If $\{a_n\}$ satisfies (2.1), then by (2.9) and Cauchy-Schwarz,

$$\begin{aligned} [P_n\{|\hat{f}(0) - f_n(0)| \leq a_n\}]^2 &\leq P_0\{|\hat{f}(0) - f_n(0)| \leq a_n\} (1 + I)^n \\ &= \{1 + o(1)\} P_0\{|\hat{f}(0) - f_n(0)| \leq a_n\}, \end{aligned}$$

so that both $P_0\{|\hat{f}(0) - f_n(0)| \leq a_n\}$ and $P_0\{|\hat{f}(0) - f_0(0)| \leq a_n\}$ converge to one as $n \rightarrow \infty$. Hence $|f_n(0) - f_0(0)| \leq 2a_n$ for large n . But $|f_n(0) - f_0(0)| = \eta b_l(2l)!/l!\pi^{\frac{1}{2}} \geq 2CB(\log n)^{-k/2}$, where C does not depend on B . Therefore $a_n \geq CB(\log n)^{-k/2}$ for large n . Since this is true for each $B > 0$ then $(\log n)^{k/2}a_n \rightarrow \infty$, completing the proof of Theorem 1.

The same construction can be used to show that if $k = 2$, the distribution function of X can be estimated at a rate no faster than $(\log n)^{-3/2}$. Let F_n and F_0 be the distribution functions for f_n and f_0 in the proof of Theorem 1, and evaluate them at ϵx_0 , where $x_0 > 0$. The calculations rely on an approximation to $H_{2l-1}(x_0)$ given by Magnus *et al.* (1966, p.254) and various integral identities on p. 251 of the same reference. We omit the details.

3. DECONVOLUTION FOR GENERAL ERRORS

There are versions of Theorems 1 and 2 for a variety of different types of error distributions. The general principle is that "the smoother the residual distribution, the slower is the optimal achievable rate of convergence". It is convenient to consider this principle in the Fourier domain, bearing in mind that smoother distributions have characteristic functions with thinner tails. If X , Y and Z have respective characteristic functions ϕ_X , ϕ_Y and ϕ_Z , and if $Y = X + Z$ where X and Z are independent, then the characteristic function of X is recoverable from that of Y via the formula $\phi_X = \phi_Y / \phi_Z$. Any data-based form of this inversion becomes increasingly difficult as the tails of ϕ_Z become thinner. For example, if Z has a gamma distribution with shape parameter α , then the tails of $\phi_Z(t)$ decrease like $|t|^{-\alpha}$ as $|t| \rightarrow \infty$, and so deconvolution is difficult for large α . In fact, the fastest achievable rate of convergence over densities in $C_k(B)$ is $n^{-k/(2k+2\alpha+1)}$. This is made clear by the following analogue of Theorem 1. Again, $\hat{f}(x_0)$ is a nonparametric estimator of $f(x_0)$.

Theorem 3. Assume that the error distribution is gamma with shape parameter $\alpha > 0$.

If, for some sequence of positive constants $\{a_n, n \geq 1\}$, we have

$$\liminf_{n \rightarrow \infty} \inf_{f \in \mathcal{C}_k(B)} P_f \{ |\hat{f}(x_0) - f(x_0)| \leq a_n \} = 1$$

for each $B > 0$, then

$$\lim_{n \rightarrow \infty} n^{k/(2k+2\alpha+1)} a_n = +\infty. \quad (3.1)$$

The “double gamma” case, where Z is symmetric and $|Z|$ is $\text{gamma}(\alpha)$, is similar. There, Theorem 3 continues to hold for integer α , provided 2α in (3.1) is changed to $4(\alpha - [\alpha/2])$, where $[\alpha/2]$ denotes the largest integer not exceeding $\alpha/2$. In particular, the optimal rate of convergence when errors have a double exponential distribution is $n^{-k/(2k+5)}$.

Proofs of results such as Theorem 3, where “algebraic” rates are available, run as follows. Let $\epsilon \rightarrow 0$ as $n \rightarrow \infty$, and fix a k -times differentiable density f_0 which is bounded away from zero in a neighbourhood of the origin. Let H be a bounded, compactly supported function with at least k bounded derivatives, and satisfying $H(0) \neq 0$ and $\int x^j H(x) dx = 0$ for $0 \leq j < \alpha + 1$. Put $f_n(x) \equiv f_0(x) + \epsilon^k H(x/\epsilon)$, and let g_n and g_0 be the convolution densities for f_0 and f_n respectively. It may be shown that if $\epsilon = n^{-1/(2k+2\alpha+1)}$ then I , defined at (2.8), satisfies $I = O(n^{-1})$. Then, arguing much as in the proof of Theorem 1, the best attainable rate of convergence emerges as being no faster than ϵ^k . Similar techniques show that for smooth, infinitely differentiable error densities such as the Cauchy, the optimal convergence rate is slower than n^{-a} for any $a > 0$.

Stefanski-Carroll (1987b) type kernel estimators achieve optimal rates in the normal, gamma and “double gamma” cases. For the sake of brevity we have omitted a proof in the latter two cases.

4. DISCUSSION

Deconvolution problems are important in their own right, as well as in nonparametric estimation of priors. In measurement error models, deconvolution arises if one wishes

to use data to suggest models for the unobservable predictors or to estimate conditional moments useful in likelihood calculations. When the measurement errors are normally distributed, our results are pessimistic, suggesting that it will be difficult to deconvolve effectively over a wide class of distributions for X .

ACKNOWLEDGEMENT

The work of R.J. Carroll was supported by the Air Force Office of Scientific Research and completed while visiting the Australian National University.

REFERENCES

- Berger, J.O. (1980). *Statistical Decision Theory*. Springer-Verlag, New York.
- Carroll, R.J., Spiegelman, C.H., Lan, K.K., Bailey, K.R. & Abbott, R.D. (1984). On errors-in-variables for binary regression models. *Biometrika* **71**, 19–25.
- Magnus, W., Oberhettinger, F. & Soni, R.P. (1966). *Formulas and Theorems for the Special Functions of Mathematical Physics*. Springer-Verlag, Berlin.
- Maritz, J.S. (1980). *Empirical Bayes Methods*. Methuen, London.
- Medgyessy, P. (1977). *Decomposition of Superpositions of Density Functions and Discrete Distributions*. John Wiley, New York.
- Mendelsohn, J. & Rice, R. (1982). Deconvolution of microfluorometric histograms with B splines. *J. Am. Stat. Assoc.* **77**, 748–53.
- Prakasa Rao, B.L.S. (1983). *Nonparametric Functional Estimation*. Academic Press, New York.
- Sansone, G. (1959). *Orthogonal Functions*. Interscience, London.
- Schafer, D.W. (1987). Covariate measurement error in generalized linear models. *Biometrika* **74**, 385–391.
- Stefanski, L.A. (1985). The effects of measurement error on parameter estimation. *Biometrika* **72**, 583–592.

- Stefanski, L.A. & Carroll, R.J. (1985). Covariate measurement error in logistic regression. *Annals of Statistics* **13**, 1335–1351.
- Stefanski, L.A. & Carroll, R.J. (1987a). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika*, to appear.
- Stefanski, L.A. & Carroll, R.J. (1987b). Deconvoluting kernel density estimators. University of North Carolina Mimeo Series #1623.

END

DATE

FILMED

DTIC

9-88